# A deep 3D residual CNN for false-positive reduction in pulmonary nodule detection

Hongsheng Jin
*State Key Lab of CAD & CG, Zhejiang University, Hangzhou 310027, China*

Zongyao Li
*The School of Aeronautics and Astronautics, Zhejiang University, Hangzhou 310027, China*

Ruofeng Tong[a)] and Lanfen Lin
*State Key Lab of CAD & CG, Zhejiang University, Hangzhou 310027, China*

(Received 19 July 2017; revised 13 February 2018; accepted for publication 17 February 2018; published 25 March 2018)

**Purpose:** The automatic detection of pulmonary nodules using CT scans improves the efficiency of lung cancer diagnosis, and false-positive reduction plays a significant role in the detection. In this paper, we focus on the false-positive reduction task and propose an effective method for this task.
**Methods:** We construct a deep 3D residual CNN (convolution neural network) to reduce false-positive nodules from candidate nodules. The proposed network is much deeper than the traditional 3D CNNs used in medical image processing. Specifically, in the network, we design a spatial pooling and cropping (SPC) layer to extract multilevel contextual information of CT data. Moreover, we employ an online hard sample selection strategy in the training process to make the network better fit hard samples (e.g., nodules with irregular shapes).
**Results:** Our method is evaluated on 888 CT scans from the dataset of the LUNA16 Challenge. The free-response receiver operating characteristic (FROC) curve shows that the proposed method achieves a high detection performance.
**Conclusions:** Our experiments confirm that our method is robust and that the SPC layer helps increase the prediction accuracy. Additionally, the proposed method can easily be extended to other 3D object detection tasks in medical image processing.  [https://doi.org/10.1002/mp.12846]

Key words: computer-aided diagnosis (CAD) system, deep learning, false positive reduction, pulmonary nodule detection, 3D residual CNN

## 1. INTRODUCTION

The early detection of lung nodules with low-dose CT scans is very important for the diagnosis and clinical management of lung cancer.[1] However, analyzing large numbers of CT scans would be a considerable burden for radiologists. Therefore, the automatic detection of pulmonary nodules plays a significant role in computer-aided detection (CAD) systems. This will considerably increase the efficiency of lung nodule screening.

Many CAD systems have already been proposed for this task. The typical setup of a CAD system includes two steps: (a) candidate nodule detection and (b) false-positive reduction. The first step is to screen the entire CT scan and recommend candidate locations where nodules are likely to be. During this step, it is vital to work with the entire CT scan efficiently and achieve a very high sensitivity. However, this step typically results in many false-positive candidates. Therefore, we have the second step, which aims to distinguish the candidates and reduce false positives. The false-positive reduction task plays a decisive role in the performance of the system. The sensitivity and false-positive rate are combined to evaluate the performance of the CAD system.

Over the past years, automatic pulmonary nodule detection methods have attracted substantial interest.[2–5] The false-positive reduction task of pulmonary nodule detection is still a challenge for several reasons. On the one hand, the pulmonary nodules show wide variations in shapes, sizes, and types,[6] as shown in Fig. 1(a). On the other hand, as shown in Fig. 1(b), some pulmonary interstitiums (e.g., blood vessels and pulmonary fibrosis) have appearances that are similar to the true pulmonary nodules, which makes the distinguishing task considerably more difficult.

Recently, deep convolutional neural networks (CNNs) have achieved great success in image processing[7–11] and have also been introduced into the medical image field. In a recent work, Setio et al.[12] proposed a 2D multiview CNN to reduce false positives for pulmonary nodule detection. This method was a state-of-the-art achievement superior to the works using handcrafted features.[3,4] It reached a sensitivity of 76.0% at 0.25 false positives per scan on the benchmark of the LIDC-IDRI dataset.[13] The multiview architecture allows the 2D CNNs to involve spatial information, but it still cannot contain enough 3D spatial information, particularly for nodules with irregular shapes. Moreover, Dou et al.[14] combined three 3D CNNs with different input sizes to extract multilevel contextual

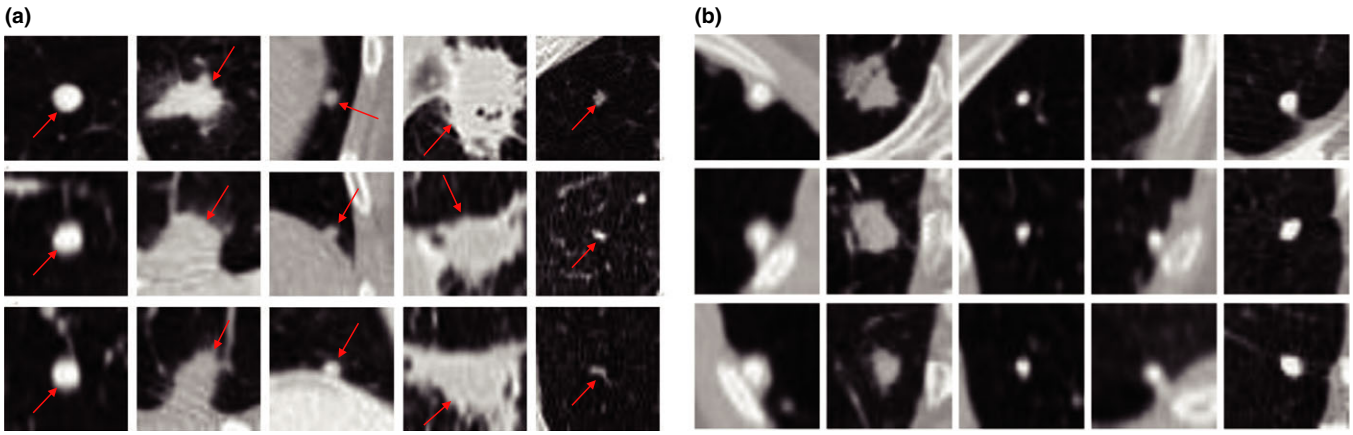


FIG. 1. The axial, sagittal, and coronal views of some pulmonary tissue examples. Subfigure (a) shows some pulmonary nodules with wide variations in shapes, sizes, and types. Subfigure (b) shows that some pulmonary interstitiums are quite similar to the true nodules. In each subfigure, the first row shows the axial view, the second row displays the sagittal view, and the third row is the coronal view. [Color figure can be viewed at wileyonlinelibrary.com]

information and achieved a sensitivity of 83.4% at 0.25 false positives per scan on the LIDC-IDRI dataset. Nevertheless, the networks that they designed are relatively shallow and do not have sufficient representative capability. Their combination method also introduced many redundant computations.

In this paper, we propose a method based on a deep 3D residual CNN with a particular spatial pooling and cropping (SPC) layer. A 2D residual network was proposed by He et al. in 2015[15] and has attracted significant research interest.[16–18] The residual networks are easier to optimize and can benefit from increased depth. To make the network involve multiscale information of patches, we design the SPC layer, which merges a max-pooling operation and a cropping operation. Compared with the ordinary residual network, our residual network with SPC layers can achieve higher detection sensitivity at the same false positives per scan, benefiting from the multilevel contextual information. Additionally, to solve the sample imbalance problem, we implement an online hard sample selection (OHSS) strategy in the training process. The strategy is inspired by the online hard example mining (OHEM) algorithm,[19] which was proposed for object detection tasks. We demonstrate that this strategy is quite effective at the false-positive reduction task in pulmonary nodule detection. Finally, in the testing process, we perform a multitest strategy to reduce the influence of overfitting.

Our contributions can be summarized as follows:

1. We propose a method for false-positive reduction using a deep 3D residual convolution network. Our network is much deeper than the networks in the previous methods[12,14] and thereby has potentially greater ability for representing features.
2. We design an SPC layer to make the network involve multilevel contextual information. Compared with other methods,[14,20] the SPC layer is more suitable for this detection task and more efficient.
3. We perform the online hard sample selection and multitest strategy to enhance the performance and prove the effectiveness of these strategies.

## 1.A. Related work

### 1.A.1. Pulmonary nodule detection

Over the past few years, several methods have been proposed for automatic pulmonary nodule detection[2–5] utilizing traditional machine learning methods. Murphy et al.[2] proposed a CAD system consisting of a candidate detector based on two successive k-nearest-neighbor classifiers with local image features of shape index and curvature for false-positive reduction. Jacobs et al.[4] trained a classifier with a set of 128 features, including intensity, shape, texture, and context features. Recently, with the great success of CNNs in image processing, some researchers in the medical image field have utilized CNNs to automatically learn features. For example, Setio et al.[12] proposed a 2D multiview CNN to reduce false positives. The network has a parallel architecture with multiple inputs. Each input is a different view of the candidate CT patch. The multiview architecture allows the 2D CNN to involve spatial information, which is necessary for extracting more representative features. Dou et al.[14] utilized 3D CNNs to address false-positive reduction. They trained three 3D CNNs with different input sizes and eventually merged the results of these networks. The inputs with different sizes contain information in different scales, and the merging of multiscale information of candidates can help the system achieve better performance than a single-scale network. However, these CNN structures are relatively shallow and cannot extract enough representative features from medical images.

### 1.A.2. 3D CNN in medical images

Many medical images, such as CT scans and MRIs, are 3D volume data. Recently, some 3D CNNs have been proposed for processing these volumetric medical images.[21–24] Kamnitsas et al.[21] presented a 3D parallel CNN for brain lesion segmentation. The parallel CNN processed input images in multiple scales. Subsequently, they used a 3D fully connected conditional random field (CRF) to reduce false

positives. Çiçek et al.[23] adapted the previous U-Net architecture for 3D applications. Dou et al.[24] proposed a CAD system composed of a 3D fully convolutional network (FCN) and a 3D CNN for detecting cerebral microbleeds from MR images. The 3D FCN could screen the MR images and recommend suspect candidates, and the 3D CNN was used to discriminate true cerebral microbleeds. These works demonstrate that 3D CNNs can extract available 3D spatial information. In our work, we further show that it is essential to design deeper 3D CNNs to extract 3D spatial information more effectively.

#### *1.A.3. Residual learning*

Recently, He et al.[15] presented the residual learning framework. They designed a shortcut connection among layers that can increase information propagation. He et al.[16] further proposed a new residual unit. The new residual unit made training deep networks easier and improved generalization. Later, many works enhanced and attempted to improve the residual framework.[17,18] Many general image detection and classification works[9,25,26] also utilized this idea and achieved promising performance. For example, Dai et al.[25] adopted a 101-layer 2D residual network as the fully convolutional image classifier backbone for their R-FCN objection detection method. The R-FCN method achieved promising results in natural object detection and achieved 83.6% mean average precision (mAP) on the VOC 2007 dataset. However, to the best of our knowledge, the effectiveness of residual learning on medical image classfication tasks has not been extensively explored.[27,28]

#### *1.A.4. Multilevel information*

In the area of general image processing, multilevel contextual information has been shown to be useful.[29] To leverage the multilevel contextual information, He et al.[20] designed a spatial pyramid pooling layer for visual recognition. However, the inception network[26] adopted another idea. They utilized convolution kernels with different sizes to extract the multilevel information. Recently, some studies have begun to consider multilevel contextual information in medical image processing. Dou et al.[14] designed three CNNs of different sizes and subsequently merged these outputs. Moeskops P. et al.[30] designed a 2D parallel CNN with multiscale inputs. Each branch of the parallel network has a different convolution kernel size. Kamnitsas et al.[21] further presented a 3D parallel CNN with multiscale inputs. The parallel structures or the combination strategy in these methods introduce many redundant computations.

#### *1.A.5. Hard example mining*

In the field of image classification and objection detection, an imbalance between object and background always exists. Additionally, sample selection strategies have been studied for a long time.[19,31–33] Rowley et al.[31] used a bootstrap algorithm to train a shallow neural network for face detection. The algorithm continuously uses this network to find false positives and adds them to the training set. Felzenszwalb et al.[32] proposed a hard example mining algorithm for training a support vector machine (SVM). They designed an algorithm that schedules the model training and the training set updating. Loshchilov et al.[33] proposed an online batch selection method to accelerate AdaDelta and the Adam algorithm. The selection strategy is ranking the samples by loss and decreasing the batch size during the training process. Recently, Shrivastava et al. proposed the online hard example mining algorithm for object detection tasks.[19] Their work was based on fast R-CNN[8] and the SGD algorithm. The strategy is to sort the input regions of interest (RoIs) by loss and train with the examples performing the worst, which leads to an improvement of 4.1 points in mAP on the VOC12 dataset. We show that the online hard example selection strategy is helpful for solving the imbalance problem in medical detection tasks.

## 2. METHOD

### 2.A. Residual learning

The residual network has been proven to be effective for general image processing tasks. In our work, we also find that it is more effective than traditional CNNs. Residual learning allows units of stacked layers to fit a residual mapping rather than a desired underlying mapping directly.[15] Formally, the original mapping can be expressed as a residual form:

$$\mathcal{F}(x) = \mathcal{H}(x) - x \tag{1}$$

where $\mathcal{H}(x)$ is the original underlying mapping function, $x$ is the input feature, and $\mathcal{F}$ is the residual function. The original function thus becomes $\mathcal{F}(x) + x$; therefore, there is an identity mapping (also called short connection) among layers, which can allow signals to be directly propagated from one residual unit to another.

To some degree, residual learning solves the vanishing gradient problem in deep neural networks. The identity mapping prevents the network from degenerating with increasing network depth. Moreover, the identity mapping further enhances and smooths the information propagation during training.[16] In this paper, we extend the residual unit into 3D form to make the training process of deep 3D CNNs more effective, even with limited image data. Our 3D residual unit is shown in Fig. 2.

### 2.B. Architecture

To fully utilize deep residual learning and the 3D spatial contextual information of volumetric images, we extend the 2D deep residual network to a 3D form. The overall architecture of our 3D deep residual convolutional network for the false-positive reduction task in pulmonary nodule detection is shown in Fig. 3(a). Specifically, we design a 27-layer network that includes three residual groups. Each residual group

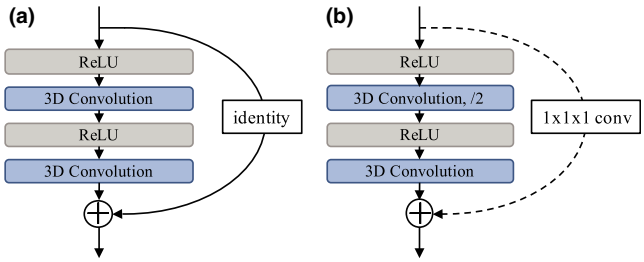


FIG. 2. 3D residual unit. The 3D residual unit (a) is the normal architecture with an identity map. The 3D residual unit (b) downsamples the input with the first convolutional layer, which is labeled as /2. The $1 \times 1 \times 1$ convolutional kernel in the 3D residual unit (b) is utilized when the size of the feature map or number of channels is changed. It performs a linear projection to match the size of the feature map and number of channels between the input and output. [Color figure can be viewed at wileyonlinelibrary.com]

contains four residual units. Each residual unit consists of two convolutional layers with a kernel size of $5 \times 5 \times 3$. In each residual unit, the input feature $x$ and the residual mapping function $\mathcal{F}(x)$ are added together; thus, the signals can transfer directly through the identity mapping. Furthermore, the first two residual groups are followed by a spatial pooling and cropping (SPC) layer. Following the design rule of Resnet,[15] we doubled the number of filters to preserve the time complexity per layer when the SPC layer is used. Rectified linear units (ReLU)[34] are used as the activation functions of all layers in the network. We also utilize a dropout layer to improve the generalization capability of the network.[35] All the layers mentioned are implemented in a 3D manner.

In addition, because 3D CNNs require massive memory resources, the size of the input data and the depth of the network are greatly limited. We fully utilize the graphics card memory to make our architecture as deep as possible. Although this neural network is still shallower and contains fewer convolutional filters than 2D residual networks applied for general image classification tasks, it has been proven to be considerably more effective than prior shallow CNNs in the false-positive reduction task of lung nodule detection through the following experiments.

## 2.C. Spatial pooling and cropping layer

Pulmonary nodules exhibit a wide variation in sizes, shapes, and types. With regard to size variations, the diameters of the nodules range from 3 to 30 mm. Additionally, the nodules are of varying types such as solid, subsolid, calcified, and pleural. In addition, the surroundings of the nodules are often complicated. To address this problem, Dou et al. proved that the receptive field size, i.e., the surrounding range of a target position, greatly influences the accuracy of the predictions.[14] Due to the large size variations and complicated surroundings of the nodules mentioned above, it is necessary to extract multilevel features via different receptive field sizes. In previous works, typical strategies for involving the multilevel contextual information of training images include (a) designing a parallel network that contains several inputs and merges the branches in one layer[20,30] and (b) training several networks with inputs at different scales and using the fusion of these models to achieve a final prediction.[14] However, there are still several shortcomings in their methods. First, both methods require much more training time than a single network with a single input. Second, it is necessary to carefully select the receptive field size of the networks and tune the fusion parameters. In this paper, the SPC layer that we proposed makes the network fuse the multilevel contextual information of image data automatically.

Figure 4 shows the 3D SPC layer. This layer consists of a pooling part and a cropping part. The pooling part is 3D max-pooling, which obtains the global contextual information of the input feature maps. The cropping part aims to crop the central volume of the 3D feature map, and the result of cropping has half the size of the feature map. The cropping part extracts the information of the central areas, which involves the nodules in the feature maps. Clearly, the results of both pooling and cropping have the same size. Therefore, the output of the SPC layer is to concatenate the results of the pooling and cropping on the dimension of the channel, and the number of channels will be doubled. General residual networks always double the number of convolutional filters when the feature map size is halved. We place the SPC layers in front of the convolutional layers, which doubles the number of filters. Consequently, although the SPC layer doubles the number of channels, the increase in the number of connections in the network is minimal, and the increased training cost is slight. Compared with previous fusion methods, the convolution layers and fully connected layers in our network are completely shared, which makes the network more efficient and effective.

Similar to our approach, SPPnet[20] can extract multilevel contextual information. It uses spatial pyramid pooling and concatenates the pooling results together prior to the fully connected layer. However, in our nodule detection task, simply downsampling the feature map at multiple scales may lose significant amounts of important contextual information. We need to better focus on the central information, particularly for small nodules. The cropping part of our SPC layer will preserve the central feature map information and eliminate the interfering surroundings.

## 2.D. Online hard sample selection

For the false-positive reduction task of lung nodule detection, we observe that most of the candidate samples are very easily fit. This means that the training loss of these easy candidate samples could easily be reduced. Nevertheless, due to the variations characterizing nodules mentioned above, several candidate samples are particularly hard to fit, and these samples are called hard candidate samples. Examples of easy candidate samples and hard candidate samples are shown in Fig. 5. There are considerably more easy samples than hard samples. When we train our network with the mini-batch stochastic gradient descent (SGD) algorithm, the loss of a mini-batch is the average loss of these samples in the mini-batch. Consequently, while the loss of the mini-batch is

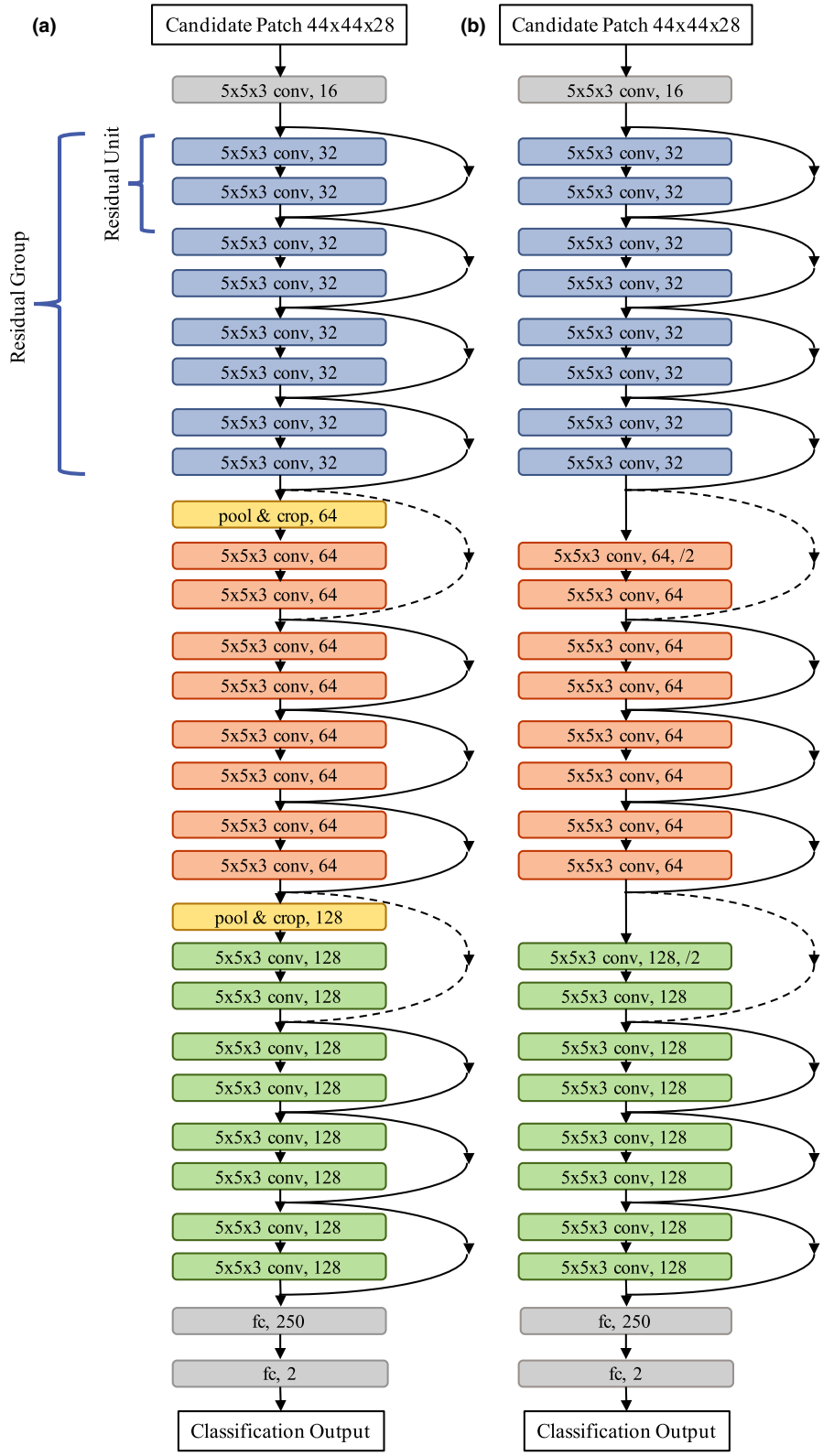


FIG. 3. The network architecture of deep 3D residual CNN. The architecture (a) is the 3D residual CNN of our method, which utilizes the SPC layer. The architecture (b) is the network for comparison, which uses the normal downsampling operation in the convolution layer as previously described.[15] The details of the residual unit with the real line shortcut are shown in Fig. 2(a). The details of the residual unit with the dashed line shortcut are shown in Fig. 2(b), where the "/2" label in the convolution layer means that the convolution stride is 2. [Color figure can be viewed at wileyonlinelibrary.com]

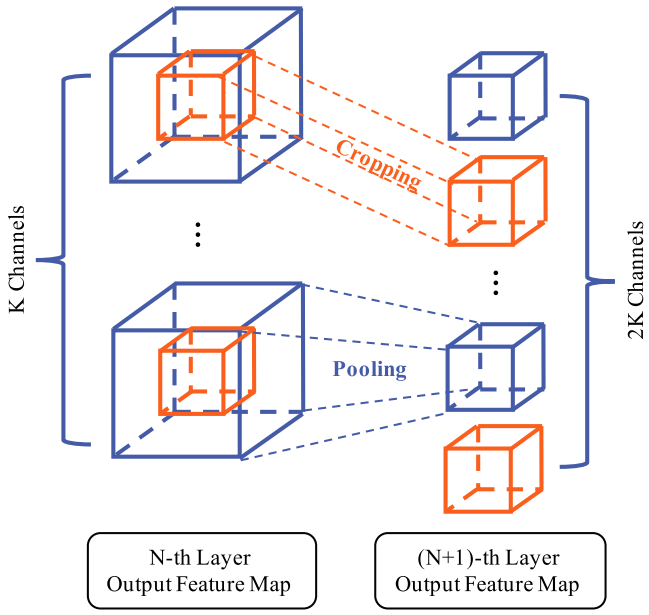


FIG. 4. The details of our designed 3D spatial pooling and cropping (SPC) layer. [Color figure can be viewed at wileyonlinelibrary.com]

relatively low, the loss of hard samples is still high. The loss could not be reduced further after only a period of training, while the network is still not trained well. It is too difficult to train the network well using the normal training strategy because the hard samples could not provide enough effects during training.

To address the problem of this extreme imbalance between easy and hard samples, we utilize the strategy of online hard sample selection. This strategy is based on the mini-batch SGD algorithm. Online hard sample selection chooses hard samples as the training data to update the weights of the network. Specifically, in each iteration, we test a mini-batch of N samples with the current network model, sort the N samples by loss and choose K hard samples that have the largest loss. The K hard samples are used to update the weights with standard back propagation[36] in each iteration. In our experiments, we set N to 128 and K to 32. In our implementation of the online hard sample selection strategy, most of the easy samples are not involved in training the network such that hard samples can provide more effects during training. Furthermore, this training strategy can reduce the number of back-propagation iterations and accelerate the training process.

Note that the strategy of online hard sample selection tends to make the training extremely difficult at the beginning of the training process. It is considerably more difficult to make the training convergent if we directly use this strategy. Therefore, we first warm up the network with the normal SGD algorithm for a few initial iterations; the training process shows a clear convergence tendency. Next, we save the model of the network that is trained normally as the initial model and continue to train the network using the strategy of online hard sample selection. Then, the training process is normally convergent.

### 2.E. Multitest

To further enhance the capabilities of our method, we utilize a multitest strategy during the testing process. First, we extend several samples from each original test sample. The extension method is to rotate the sample 0°, 90°, 180°, or 270° in the x-y plane and translate the sample 0, −1 or 1 pixels randomly in each of three dimensions. Next, we evaluate the extended test samples and calculate the average prediction probability of these extended samples as the probability of the test sample. The multitest strategy plays the role of reducing the risk of overfitting and is able to considerably enhance the final performance.

## 3. EXPERIMENTAL

### 3.A. Dataset and data processing

The dataset used to train and evaluate our method is from the LUNA16 Challenge.[37] The dataset is based on the publicly available LIDC/IDRI database, excluding the scans with a slice thickness greater than 2.5 mm. In total, there are 888 CT scans included in this dataset. Every scan has been annotated by four experienced radiologists. Each radiologist marked the lesions that they identified as non-nodule, nodule <3 mm and nodule ≥3 mm. All nodules ≥3 mm accepted by at least three out of four radiologists are selected as the reference standard. The total number of nodules in this dataset is 1186. Non-nodules and the remaining nodules are referred to as irrelevant findings and are ignored during the evaluation.

In the LIDC/IDRI dataset, these 888 CT scans are acquired from different hospitals using different equipment. Furthermore, the variations in these CT scans in terms of spacing and dynamic range are very large. The spacing ranges from 0.46 × 0.46 × 0.45 mm to 0.97 × 0.97 × 2.50 mm. We employ a simple preprocessing strategy to reduce these variations. First, we unify the spacing of each scan as 0.75°, 0.75°, and 1.25 mm in the X, Y, and Z directions with 3D linear interpolation. Then, we clip the Hounsfield unit (HU) values of the CT scans into the interval (−1000, 400) and normalize them into the range of (0, 1).

Because the number of true positive candidates is substantially smaller than the total number of candidates (almost 490:1), data augmentation must be performed. We employ a simple data augmentation strategy to enlarge the true positive candidate sample set. We crop a 48 × 48 × 28 sub-volume around the candidate location as a sample patch, and then, we extend 108 samples from each true positive sample by combining the following two operations: (a) rotating the sample 0°, 90°, 180°, or 270° in the x-y section and (b) translating the sample −1, 0, or 1 pixels in each of three dimensions.

### 3.B. Implementation

Our 3D deep residual CNN is implemented based on the deep learning library Keras. The training and evaluation processes of this network are performed on a NVIDIA TITAN X

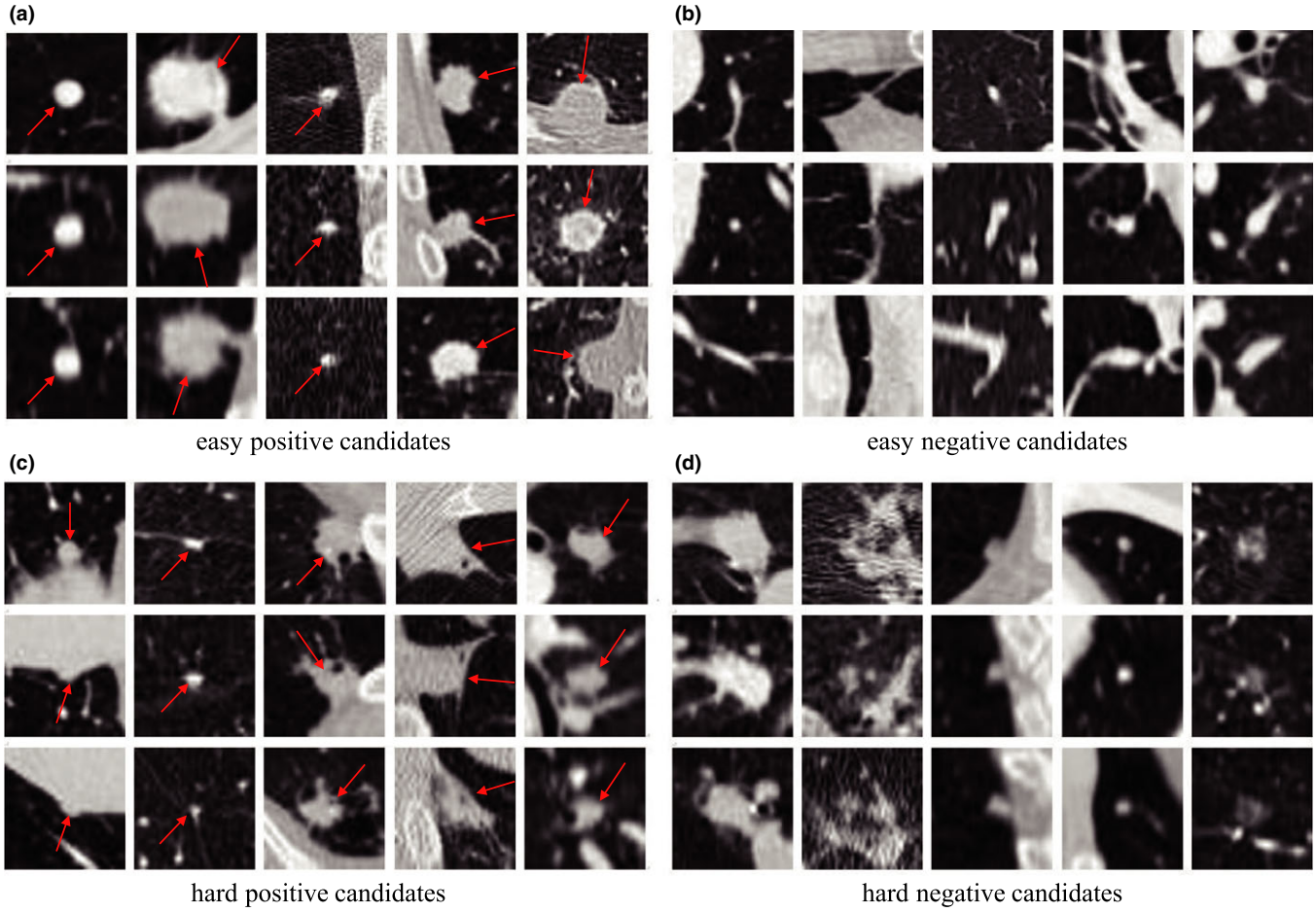


FIG. 5. Examples of pulmonary nodule candidates in axial, sagittal, and coronal views. Candidates (a) and (c) are true nodules. Candidates (b) and (d) are non-nodules. Furthermore, candidates (a) and (b) can be easily classified, and candidates (c) and (d) are difficult to be classified. [Color figure can be viewed at wileyonlinelibrary.com]

GPU. In the training phase, we initialize the weights of our network with MSRA weight[38] and warm up the network with a normal mini-batch stochastic gradient descent for 1500 iterations. Then, we train the network with the online hard sample selection strategy for 16000 iterations. In the online hard sample selection strategy, due to the limitations imposed by GPU memory, we employ a hard example size K of 32 and mini-batch size N of 128. The initial learning rate is set as 0.01, and it is tuned to 0.001 after 8000 iterations. In the test phase, we set the extended number of each sample as 20, and more extended samples do not further enhance the performance.

### 3.C. Results and comparison

To validate the performance of our method, we evaluate our method on the LUNA16 Challenge. Because the ground truth of the entire dataset was provided by the organizers, we perform tenfold cross-validation on the provided dataset and obtain the final result. Moreover, to perform comparative experiments with other false-positive reduction methods, we use the nodule candidate list (NCL) provided by the sponsor of the challenge. In this candidate list, 551,065 nodule candidates are computed using three candidate detection algorithms.[2,4,5] Within these candidates, 1120 out of 1186 nodules are detected, which means that the highest sensitivity that we can reach is 94.4%. The ratio of positive and negative candidates in this candidate list is 1:407. Recently, the sponsor provided a new candidate list, in which 1166 out of 1186 nodules are detected, and the highest sensitivity that we can reach is 98.3%. In this new candidate list, the ratio of positive to negative candidates is 1:483. We also evaluate our method on this new candidate list. In the following part, for simplicity, we denote the previous candidate list as V1 and the later one as V2.

The evaluation is performed by measuring the detection sensitivity and the corresponding false-positive rate per scan. We need to submit the raw prediction probabilities of nodule candidates to the LUNA16 Challenge; then, the free-response receiver operating characteristic (FROC) analysis is performed by the organizers.[1] The FROC curve of our method is shown in Fig. 6. The quantitative score in the challenge is defined as the average sensitivity at seven predefined false-positive rates: 1/8, 1/4, 1/2, 1, 2, 4, and 8 false positives (FPs) per scan. The quantitative score of our method is shown in Table I.

We compare our results with state-of-the-art methods that also participated in the LUNA16 Challenge. Additionally, the quantitative scores of these methods are listed in Table I

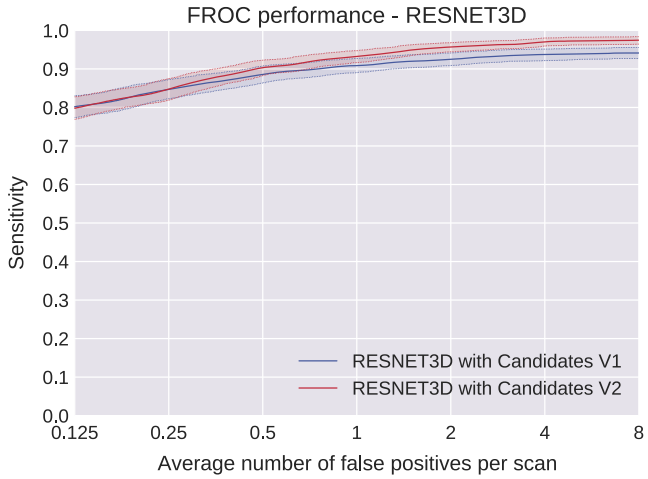


FIG. 6. The FROC curve of the results of our method. Dashed lines denote the 95% confidence interval estimated via bootstrapping.

according to the order of submission date. Before our submission, DIAG CONVNET and CUMedVis are the best two methods. The DIAG CONVNET method proposed by Setio et al.[12] uses a 2D multiview convolutional network. The CUMedVis method proposed by Dou et al.[14] trains three 3D CNNs and merges the results of the three networks as the final result. 3D CNNs can extract 3D spatial contextual information of volumetric images, but the shallow architecture limits the representative capability of CNNs.

Compared with these methods, our architecture achieves higher sensitivity. Specifically, when the number of false positives per scan is low, the sensitivity that we reached is higher than that of previous methods by almost 10%.

Our method is more effective than the previous methods as a result of the following advantages. Compared with the DIAG CONVNET method, the 3D structure and SPC layer make our network involve richer spatial and contextual information. Compared with the CUMedVis method, the residual learning framework that we utilize allows us to design much deeper 3D CNNs for obtaining more representative features. After our submission, many other teams submit their new results. Among them, some teams like JianPeiCAD, qfpxfd, MILAB, IHPC, LUNA16FONOVA-CAD, and PATech achieve better detection performance. However, we are unable to conduct further discussions, since they have not published papers yet and only provide limited descriptions. In the future, we are looking forward to more communications with these teams and further enhancing our method.

## 3.D. Ablation experiments

We conduct a number of ablation experiments to analyze our proposed method. In the ablation experiments, the dataset is split into ten subsets as in previous work.[12,14] We train on subsets 1–8, validate on subset 9, and test on subset 0. The results are shown in Table II and discussed in detail in the following contexts.

### *3.D.1. Spatial pooling and cropping layer*

We design a network without SPC layers for comparison to validate the effectiveness of SPC layers through a comparative experiment. The compared network is kept similar to the original residual network architecture.[15] The SPC layers in this network are removed, and the number of strides of the convolutional layers following the SPC layers is set as 2, as shown in Fig. 3(b). Table II further shows the quantitative performance comparison of the 3D deep residual network with and without an SPC layer. The network with an SPC layer can achieve better performance. This demonstrates that an SPC layer can indeed extract multilevel contextual information around pulmonary nodules and benefit the pulmonary nodule detection.

TABLE I. The quantitative results of the proposed method and other methods.

| Team | Date | NCL | 0.125 | 0.25 | 0.5 | 1 | 2 | 4 | 8 | Score |
|---|---|---|---|---|---|---|---|---|---|---|
| CUMedVis (QiDou)[14] | 2 April 2016 | V1 | 0.678 | 0.738 | 0.816 | 0.838 | 0.879 | 0.907 | 0.922 | 0.827 |
| DIAG_CONVNET (arnaud.setio)[12] | 1 May 2016 | V1 | 0.692 | 0.771 | 0.809 | 0.863 | 0.895 | 0.914 | 0.923 | 0.838 |
| JianPeiCAD (weiyixie) | 24 December 2016 | V1 | 0.743 | 0.826 | 0.896 | 0.931 | 0.940 | 0.943 | 0.944 | 0.889 |
| **RESNET3D (Ours)** | 17 May 2017 | V1 | 0.802 | 0.847 | 0.886 | 0.909 | 0.925 | 0.936 | 0.941 | **0.892** |
| MILAB_ResCAD (bckim) | 21 July 2017 | V1 | 0.767 | 0.879 | 0.901 | 0.911 | 0.917 | 0.921 | 0.929 | 0.889 |
| CUMedVis (QiDou)[14] | 15 May 2016 | V2 | 0.677 | 0.834 | 0.927 | 0.972 | 0.981 | 0.983 | 0.983 | 0.908 |
| DIAG_CONVNET (arnaud.setio)[12] | 23 May 2016 | V2 | 0.669 | 0.760 | 0.831 | 0.892 | 0.923 | 0.944 | 0.960 | 0.854 |
| **RESNET3D (Ours)** | 17 May 2017 | V2 | 0.794 | 0.847 | 0.904 | 0.933 | 0.957 | 0.971 | 0.975 | **0.912** |
| qfpxfd (pku.hzq) | 26 May 2017 | V2 | 0.797 | 0.857 | 0.895 | 0.938 | 0.954 | 0.970 | 0.981 | 0.913 |
| JianPeiCAD (weiyixie) | 21 July 2017 | V2 | 0.725 | 0.859 | 0.922 | 0.963 | 0.980 | 0.982 | 0.983 | 0.916 |
| IHPC_zkj (zkj) | 24 July 2017 | V2 | 0.874 | 0.966 | 0.981 | 0.983 | 0.983 | 0.983 | 0.983 | 0.965 |
| MILAB_ConcatCAD (bckim) | 21 August 2017 | V2 | 0.906 | 0.928 | 0.935 | 0.939 | 0.949 | 0.957 | 0.963 | 0.940 |
| LUNA16FONOVACAD (zxp774747) | 15 September 2017 | V2 | 0.945 | 0.957 | 0.965 | 0.970 | 0.974 | 0.976 | 0.978 | 0.966 |
| PATech (PA_tech) | 20 December 2017 | V2 | 0.919 | 0.963 | 0.973 | 0.979 | 0.981 | 0.981 | 0.981 | **0.968** |

The performance scores of the best method and our method are shown in bold.

TABLE II. The quantitative performance comparison of our network and ablated networks.

| FPs/scan | 0.125 | 0.25 | 0.5 | 1 | 2 | 4 | 8 | Score |
|---|---|---|---|---|---|---|---|---|
| RESNET3D | 0.823 | 0.877 | 0.905 | 0.931 | 0.962 | 0.979 | 0.991 | 0.924 |
| RESNET3D without SPC | 0.778 | 0.838 | 0.896 | 0.918 | 0.959 | 0.977 | 0.983 | 0.907 |
| RESNET3D without OHSS | 0.766 | 0.827 | 0.886 | 0.925 | 0.955 | 0.969 | 0.976 | 0.893 |
| RESNET3D without multitest | 0.737 | 0.826 | 0.876 | 0.925 | 0.962 | 0.979 | 0.983 | 0.898 |

#### 3.D.2. Online hard sample selection strategy

For this experiment, we train a compared model with a normal mini-batch stochastic gradient descent algorithm. From the comparison between our method with the OHSS strategy and without the strategy in Table II, we observe that the OHSS strategy helps to detect more pulmonary nodules, even at high false positives per scan. This result further shows that the OHSS strategy be better trained on hard nodule samples.

#### 3.D.3. Multitest

Table II shows the result of our method with the multitest strategy and without the multitest strategy. Our method with the multitest strategy achieves higher sensitivity. As mentioned previously, due to the lack of pulmonary nodule data, we perform heavy augmentation on the pulmonary nodule samples. This strategy can make the prediction of pulmonary nodules more stable despite the heavy augmentation.

## 4. DISCUSSION

The false positive reduction task is a critical part of CAD systems for pulmonary nodule detection and influences the performance of the entire system. In this study, we propose an effective false-positive reduction method based on a deep 3D residual CNN. Our method achieves an average sensitivity of 91.2% at seven predefined false-positive rates mentioned above, which outperforms the previous methods.[12,14] As shown in Fig. 7, our method can even correctly predict some pulmonary nodules that are difficult to detect. Some of these pulmonary nodules are extremely small, and some of them have irregular shapes. A major factor in the successful detection of extremely small nodules is the SPC layer. Information of the candidate patches on multiple scales tends to make the final prediction more accurate. Our network is able to involve information on multiple scales with the SPC layer, and the SPC layer only introduces a slight computation overhead for training and evaluating our model. The OHSS strategy also contributes to the detection of the extremely small nodules and the nodules with irregular shapes. With the OHSS strategy, these pulmonary nodules that are difficult to detect can provide enough effects in the training process. Then, our model can learn the morphological characteristics of these hard pulmonary nodules well.

Moreover, our deep 3D residual CNN greatly benefits from the 3D residual unit. The residual network can benefit from increased depth and is easier to optimize.[15] We also demonstrate that a deep 3D CNN can achieve higher performance compared to previous shallower methods.[14] However, the deep 3D CNN structure that we designed consumes a considerable amount of memory and computing resources. The large requirements on these resources greatly limit the depth of our network. Thus, in our future work, we would like to reduce the computation and

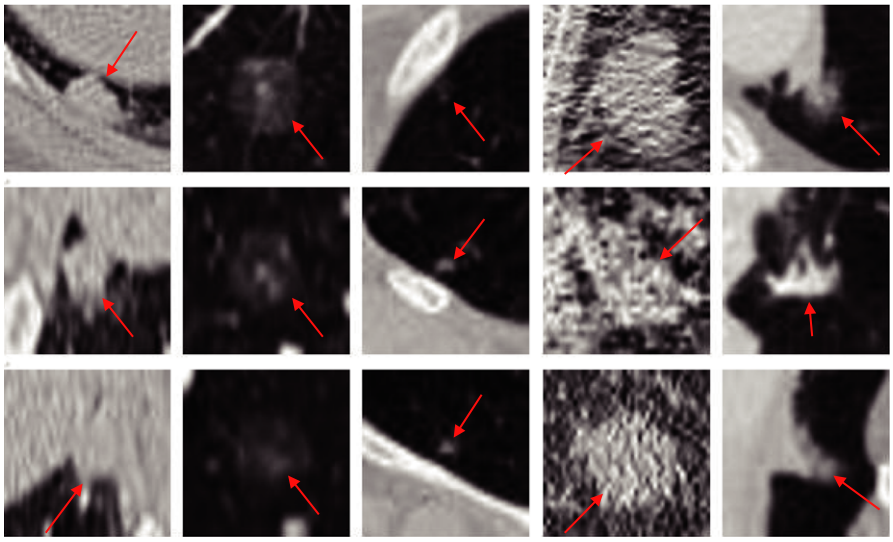

FIG. 7. Examples of pulmonary nodules that are difficult to find but that our method can correctly predict. The prediction probabilities of all these examples are greater than 0.9. [Color figure can be viewed at wileyonlinelibrary.com]

memory usage of our method without degrading the detection performance.

A typical pulmonary nodule detection CAD system has a candidate detector and a false-positive reducer. The method that we proposed is an effective false positive reducer that can address extreme imbalances between positive and negative candidate samples (1:483). Thus, we believe that a new pulmonary nodule detection CAD system that combines our method and any other candidate detector can also perform well. In other words, the detection performance of the entire system will greatly benefit from the future performance improvement of the candidate detector. In the future, we will focus on developing a candidate detector with higher sensitivity.

The LIDC/IDRI dataset used in our method is from the LUNA16 Challenge, which is the largest publicly available CT pulmonary nodule dataset that we can obtain. There are a total of 888 CT scans in this dataset. Compared with natural image classification and detection datasets, such as ImageNet and COCO, the LIDC/IDRI dataset is relatively small. We think we can try more reasonable data augmentation methods, such as reflection and zooming, in the future. Moreover, we are eager to build a larger dataset for the pulmonary nodule detection task. When the larger dataset is ready, we believe that the heavy augmentation step and the preprocessing step in our method may not be necessary.

## 5. CONCLUSION

We present a deep 3D residual CNN for the false-positive reduction task of pulmonary nodule detection in CT scans. Despite the wide variations in the shapes, sizes, and types of the pulmonary nodules, our method can achieve promising performance on the LUNA16 challenge. We show that a deep 3D CNN can achieve high nodule detection sensitivity even at a 0.25 false positives per scan rate compared to previous shallower methods. We also demonstrate that the SPC layer is helpful for extracting multilevel contextual information. The online hard sample selection strategy is effective when an extreme imbalance of easy and hard samples exists. Even at eight false positives per scan rate, the OHSS strategy helps to increase the sensitivity by 1.5%. Our method could be combined with any candidate detector that screens CT scans and recommends candidates. In the future, we wish to extend the proposed method to other medical classification tasks, particularly in volumetric medical image analysis. In addition, we would like to reduce the computational complexity of our deep 3D residual CNN method to work with volumetric medical images more efficiently.

## ACKNOWLEDGMENTS

We thank the organizers of the LUNA16 challenge for providing the dataset and evaluating our results. Moreover, we thank Dr. Qi Dou for discussing the details of the CUMedVis method with us.

## CONFLICTS OF INTEREST

The authors have no conflicts to disclose.

[a)]Author to whom correspondence should be addressed. Electronic mail: trf@zju.edu.cn

## REFERENCES

1. Setio AAA, Traverso A, deBel T, et al., Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge. arXiv preprint arXiv:1612.08012 2016.
2. Murphy K, van Ginneken B, Schilham AM, de Hoop BJ, Gietema HA, Prokop M. A large- scale evaluation of automatic pulmonary nodule detection in chest ct using local image features and k-nearest-neighbour classification. *Med Image Anal*. 2009;13:757–770.
3. Messay T, Hardie RC, Rogers SK. A new computationally efficient cad system for pulmonary nodule detection in ct imagery. *Med Image Anal*. 2010;14:390–406.
4. Jacobs C, van Rikxoort EM, Twellmann T, et al. Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images. *Med Image Anal*. 2014;18:374–384.
5. Setio AAA, Jacobs C, Gelderblom J, Ginneken B. Automatic detection of large pulmonary solid nodules in thoracic ct images. *Med Phys*. 2015;42:5642–5653.
6. Firmino M, Morais AH, Mendoça RM, Dantas MR, Hekis HR, Valentim R. Computer-aided detection system for lung cancer in computed tomography scans: review and future prospects. *Biomed Eng Online*. 2014;13:41.
7. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 2014.
8. Girshick R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision 2015:1440–1448.
9. Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector. In European Conference on Computer Vision. Springer, 2016:21–37.
10. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015:3431–3440.
11. Song Y, Li Q, Feng D, Cai W, Zou JJ. Texture image classification with discriminative neural networks. *Comput Vis Media*. 2016;2:367–377.
12. Setio AAA, Ciompi F, Litjens G, et al. Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks. *IEEE Trans Med Imaging*. 2016;35:1160–1169.
13. Armato SG, McLennan G, Bidaut L, et al. The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. *Med Phys*. 2011;38:915–931.
14. Dou Q, Chen H, Yu L, Qin J, Heng PA. Multi-level contextual 3d cnns for false positive reduction in pulmonary nodule detection. *IEEE Trans Biomed Eng*. 2016;64:1558–1567.
15. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:770–778.
16. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In European Conference on Computer Vision. Springer, 2016:630–645.
17. Targ S, Almeida D, Lyman K. Resnet in resnet: generalizing residual architectures. arXiv preprint arXiv:1603.08029 2016.
18. Veit A, Wilber M, Belongie S. Residual networks are exponential ensembles of relatively shallow networks. URL http://arxiv. org/abs/1605.06431. 2016.
19. Shrivastava A, Gupta A, Girshick R. Training region-based object detectors with online hard example mining. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016:761–769.
20. He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. In European Conference on Computer Vision. Springer, 2014:346–361.

21. Kamnitsas K, Ledig C, Newcombe VFJ, et al. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Med Image Anal*. 2017;36:61–78.
22. Dou Q, Chen H, Jin Y, Yu L, Qin J, Heng P-A. 3d deeply supervised network for automatic liver segmentation from ct volumes. In International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2016:149–157.
23. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3d u-net: learning dense volumetric segmentation from sparse annotation. In International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2016:424–432.
24. Dou Q, Chen H, Lequan Y, et al. Automatic detection of cerebral microbleeds from mr images via 3d convolutional neural networks. *IEEE Trans Med Imaging*. 2016;35:1182–1195.
25. Li Y, He K, Sun J, et al. R-fcn: Object detection via region-based fully convolutional networks. In Advances in Neural Information Processing Systems 2016:379–387.
26. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-resnet and the impact of residual connections on learning. arXiv preprint arXiv:1602.07261 2016.
27. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115–118.
28. Chen H, Dou Q, Yu L, Heng P-A. Voxresnet: deep voxelwise residual networks for volumetric brain segmentation. arXiv preprint arXiv: 1608.05895 2016.
29. Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In Computer vision and pattern recognition, 2006 IEEE computer society conference on, Vol. 2 (IEEE) 2006:2169–2178.
30. Moeskops P, Viergever MA, Mendrik AM, de Vries LS, Benders MJNL, Išgum I. Automatic segmentation of mr brain images with a convolutional neural network. *IEEE Trans Med Imaging*. 2016;35:1252–1261.
31. Rowley HA, Baluja S, Kanade T. Neural network-based face detection. *IEEE Trans Pattern Anal Mach Intell*. 1998;20:23–38.
32. Felzenszwalb PF, Girshick RB, McAllester D, Ramanan D. Object detection with discriminatively trained part-based models. *IEEE Trans Pattern Anal Mach Intell*. 2010;32:1627–1645.
33. Loshchilov I, Hutter F. Online batch selection for faster training of neural networks. arXiv preprint arXiv:1511.06343 2015.
34. Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. In Aistats, Vol. 15 2011:275.
35. Srivastava N, Hinton GE, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15:1929–1958.
36. Hinton HE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. *Neural Comput*. 2006;18:1527–1554.
37. Lung nodule analysis 2016, https://luna16.grand-challenge.org 2016.
38. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision 2015:1026–1034.